

# Performances of Various Order Selection Criteria for Autoregressive Process

Venus Khim-Sen Liew

Faculty of Economics and Business, Universiti Malaysia Sarawak, 94300 Kota Samarahan,  
Sarawak, Malaysia  
Email: [ksliew@unimas.my](mailto:ksliew@unimas.my)

To Link this Article: <http://dx.doi.org/10.6007/IJAREMS/v10-i3/10448>

DOI:10.6007/IJAREMS/v10-i3/10448

*Published Online:* 24 July 2021

## Abstract

In most economic researches, the selection of autoregressive order based for an economic time series is an essential task. Specifically, many econometric testing procedures, for instance, all forms of linearity, unit root, cointegration and causality tests, require the determination of optimal lag length selection in the first place. This study investigates the performances of various order selection criteria in selecting order of autoregressive (AR) process via a simulation study. Some 1000 independent time series for each AR process of known orders are first simulated and then subjected to lag length selection using various order selection criteria. The major findings of this study are as follows: First, the performance of various criteria in correctly estimated the true AR order deteriorates as the order grows. Second, the performance of various criteria in correctly estimated the true AR order improves as sample size grows. Third, Akaike's information criterion family (AICC, AIC) and final prediction error (FPE) are superior to other criteria for sample of size not exceeding 150 observations. Fourth, Hannan-Quinn criterion (HQC) performs better than others for sample size larger than 150 observations. Fifth, Schwarz information criterion (SIC), and Bayesian information criterion (BIC) could be useful in cases whereby a parsimony order, rather than true order is of interest; while Akaike's information criterion (AIC) and final prediction error (FPE) are better options to avoid autocorrelation in our ultimate results.

**Keywords:** Autoregressive Process, Autoregressive Order, Order Selection Criteria, Simulation

## Introduction

Autoregressive (AR) process is a very simple but useful process in the study of the behaviour of a time series. A time series is said to be generated by an AR process if its current value is a function of its own lagged (or past) values. AR process is widely applied in economic studies as most of them are conducted using time series data. It is worth pointing out that the very first thing in the application of AR process is the determination of AR order. In this respect, an AR process of order  $p$  refers to a time series in which its current value is dependent on its first  $p$  lagged values and is normally denoted by AR ( $p$ ). However, in real observations, the AR order  $p$  is always unknown and therefore has to be estimated somehow. Various order selection criteria have been proposed for this purpose. These include the Akaike's information criterion

(AIC), Akaike's bias-corrected information criterion (AICC) (also known as Modified AIC), Schwarz information criterion (SIC), Hannan-Quinn criterion (HQC), final prediction error (FPE) and Bayesian information criterion (BIC).

These criteria have been overwhelmingly adopted in time series studies. They include, just to name but limited cases, Sarantis (1999, 2001); Baharumshah *et al* (2007) that use the AIC to determine the AR order of the linearity test. Ahmed (2000); Caliñgo *et al* (2017) use the AIC and BIC and Tan and Baharumshah (1999) employ the FPE in the selection of AR order for the Granger causality test. Baum *et al* (2001); Lim and Liew (2005); Fainstein and Novikov (2011); Pickson *et al* (2017) adopt the AIC and Yamada (2000) uses AIC and HQC to determine the vector AR order of the Johansen cointegration test, while Seo and Kim (2020) adopt AIC, FPE, and HQ in Vector Error Correction (VEC) modelling. Tiwari (2011) adopts AIC to determine the optimal lag for structural Vector Autoregressive (SVAR) Model. Xu (2003) and Yang *at al*. (2014) utilize the SIC to determine the AR order for the augmented Dickey-Fuller unit root test; Liew *et al* (2004) adopts the AICC to determine the lag order for the linearity test; Hundie (2014) and Leong *et al*. (2018) employ SBC and AIC respectively for the lag section of Autoregressive Distributive Lag (ARDL) model. Liew *et al* (2012); Olaniyi and Alenoghena (2017) adopts AIC for panel unit root test, while Hatmanu (2020) adopts AICC for Ng and Perron (2001) unit root test.

Despite their popularity, practitioners are sometimes annoyed by the fact that these criteria do not always produce consistent results in the selection of AR order<sup>1</sup>. Another well-known built-in shortcoming of these criteria is that they tend to under estimate the true AR order (Guerra, 2001). Attempting to shed light on these reservations on the usage of these criteria, this study has taken the initiative to conduct a simulation study on the empirical performance of these order selection criteria. Note that the dissemination of Liew (2004) that recommends AIC and FPE for the estimation the autoregressive lag length has attracted much attention of researchers are on the use of AIC. This study includes the modified version of AIC, that is AICC to evaluate its performance over AIC and other criteria. Moreover, this study employs various lag orders, while Liew (2004) included AR (4) in the simulation. Specifically, it is interested to know (i) the top performers in terms of their capability in correctly identified the true AR order; and (ii) to what extend these criteria under estimate the true AR order in practice, if any.

The rest of this paper is organized as follows. Section 2 briefly describes the AR process, the order selection criteria and simulation procedure. Section 3 presents and discusses the results of this simulation study. Section 4 offers some concluding remarks.

## Methodology

### *Autoregressive Process*

Mathematically, an AR( $p$ ) process of a series  $y_t$  may be represented by

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t \quad (1)$$

where  $a_0$  is the intercept term and  $a_1, a_2, \dots, a_p$  are autoregressive parameters and  $\varepsilon_t$  are normally distributed random error terms with a zero mean and a finite variance  $\sigma^2$ .

The estimation of AR ( $p$ ) process involves 2 stages: First, identify the AR order  $p$  based on certain rules such as order selection criteria. Second, estimate the numerical values for intercept and parameters using regression analysis. This study is confined to the study of the performance of various commonly used order selection criteria in identifying the true order  $p$ . In particular, this study generates AR processes with  $p$  ranging from 1 to 12 and uses these criteria to determine the order of each generated series as if the order is unknown<sup>3</sup>. The autoregressive parameters are independently generated from uniform distribution with values ranging from -1 to 1 exclusively. Measures are taken to ensure that the sum of these simulated autoregressive parameters is less than unity in absolute value ( $|\sum_{i=1}^p a_i| < 1$ ) so as to avoid non-stationary AR process. The error term is generated from standard normal distribution, whereas the intercept term is omitted without loss of generality. We simulate data sets for various usable sample sizes,  $T$ : 30, 60, 120, 240, 480 and 960.<sup>4</sup> For each combination of processes and sample sizes, we simulated 1000 independent series for the purpose of order estimation<sup>5</sup>. In every case, the initial value,  $y_0$  is arbitrary set to 0.0. In our effort to minimize the initial effect, we simulate  $3T$  observations and discard the first  $2T$  observations, leaving the last  $T$  observations for order estimation. The estimated order  $\hat{p}$  is allowed to be determined from any integer ranging from 1 to 20 inclusively. In this respect, we compute the values for all 20 orders for each specific criterion and  $\hat{p}$  is taken from the one that minimizes that criterion. Note that each criterion independently selects one  $\hat{p}$  for the same simulated series.

### Order Selection Criteria

The order selection criteria to be evaluated include:

(a) Akaike information criterion,

$$AIC_p = -2T [\ln(\hat{\sigma}_p^2)] + 2p; \quad (2)$$

(b) Akaike's information corrected criterion,

$$AICC_p = -2T [\ln(\hat{\sigma}_p^2)] + 2Tp / (T - p) \quad (3)$$

(c) Schwarz information criterion,

$$SIC_p = \ln(\hat{\sigma}_p^2) + [p \ln(T)]/T; \quad (4)$$

(d) Hannan-Quinn criterion,

$$HQC_p = \ln(\hat{\sigma}_p^2) + 2T^{-1} p \ln[\ln(T)]; \quad (5)$$

(e) the final prediction error,

$$FPE_p = \hat{\sigma}_p^2 (T - p)^{-1} (T + p) \text{ and} \quad (6)$$

(f) Bayesian information criterion,

$$BIC_p = (T-p) \ln[(T-p)^{-1} T \hat{\sigma}_p^2] + T[1 + \ln(\sqrt{2\pi})] + p \ln[p^{-1} (\sum_{t=1}^T X_t^2 - T \hat{\sigma}_p^2)], \quad (7)$$

where  $\hat{\sigma}_p^2 = (T-p-1)^{-1} \sum_{t=p}^T \hat{\varepsilon}_t^2$ ,  $\varepsilon_t$  is the model's residuals and  $T$  is the sample size. Note that

the cap sign (^) indicates an estimated value. Liew (2000) provides an overview on these criteria, whereas details are given in, for instance, Brockwell and Davis (1996) and the references therein.

The main task of this study is to compute the probability of each of these criteria in correctly estimated the true autoregressive order. Note that this probability takes a value between zero and one inclusively, with a probability of zero, on one end, means that the criterion fails to pick up any true order and thereby is a poor criterion. On the other extreme, a probability of one implies that the criterion manages to correctly select the true order in all cases and hence is an excellence criterion.

Besides, we also inspect the selected orders of the estimated order for 1000 simulated series of each known order, so as to gain deeper understanding on the performance of various criteria. We will refer to the situation whereby a criterion selected lower orders than the true ones as under estimate, whereas over estimate would mean the selection of higher orders than the true ones.

### **Simulation Procedure**

Briefly, the simulation procedure involves three sub-routines: with the first sub-routine generating a series of from the AR process, whereas the second sub-routine selects the autoregressive order of the simulated series and the third sub-routine evaluate the performance of the order selection criteria. The algorithm for the simulation procedure for each combination of sample size  $T$  and AR order  $p$  is outlined as follows

1. Independently generate  $a_i$  ( $i=1,2,\dots,p$ ) from a uniform distribution in the range  $(-1, 1)$ , conditioned on  $|\sum_{i=1}^p a_i| < 1$ .
2. Generate a series of size  $3T$  from the STAR process as represented in Equation (1) of order  $p=3$  with  $a_0=0.0$  and  $a_i$  ( $i=1,2,\dots,p$ ) obtained from Step 1. Initialize the starting value,  $y_0=0.0$ . Discard the first  $2T$  observations to minimize the effect of initial value.
3. Use each selection criterion to determine the autoregressive order ( $\hat{p}$ ) for the last  $T$  observations of the series simulated in Step 2. Six selection criteria are involved.
4. Repeat Step 1 to Step 3 for  $B$  times, where  $B$  is fixed at 1000 in this study.
5. Compute the probabilities of (i) correct estimate, which is computed as  $\#(\hat{p}=P)/B$ ; (ii) under estimate, which is computed as  $\#(\hat{p}<P)/B$ ; and (iii) over estimate, which is computed as  $\#(\hat{p}>P)/B$ , where  $\#(\bullet)$  denotes numbers of time event ( $\bullet$ ) happens.

### **Results and Discussions**

The probability of various criteria in correctly identified the true order of the AR process is tabulated in Table 1. Basically, two principle findings are portrayed in this table. First, holding sample size constant, the higher the AR order, the less accurate is the criteria in picking up

the true order. For instance, with a sample size of 30, the AICC managed to correctly identify the true order of with a probability of 0.625, 0.560, 0.525, 0.433, 0.342 and 0.314 respectively for true orders 2, 4, 6, 8, 10 and 12. Similar deterioration in performance is also observed in other sample sizes for the same criterion. Indeed, in line with AICC, all other criteria also exhibit similar deterioration in performance. This upshot, which is more clearly shown in Figure 1, suggests that as the AR order grows, finding the right order will be more difficult for any order selection criterion<sup>6</sup>.

Second, holding the AR order constant, the larger the sample size, the more accurate is the criteria in picking up the true order. For example, in the case of  $p = 2$ , the probability of SIC in correctly estimated this order is 0.622, 0.626, 0.683, 0.768, 0.808 and 0.842 for  $s$  equals 30, 60, 120, 240, 480 and 960 respectively. Similar improvement in performance is also observed in other autoregressive order for the same criterion as well as for other criteria; see Figure 2 for a clearer view.

Table 1

*Probability of correctly estimated the true order of AR process, ( $\hat{p} = p$ ).*

True Order $p$	Sample Size (Logarithmic Scale)	Order Selection Criteria					
		AICC	AIC	SIC	FPE	HQC	BIC
2	30 (1.48)	0.625	0.625	0.622	0.625	0.640	0.618
	60 (1.78)	0.633	0.632	0.626	0.632	0.654	0.629
	120 (2.08)	0.696	0.696	0.683	0.696	0.712	0.696
	240 (2.38)	0.745	0.745	0.768	0.745	0.785	0.769
	480 (2.68)	0.755	0.755	0.808	0.755	0.804	0.816
	960 (2.98)	0.801	0.801	0.842	0.801	0.852	0.849
4	30 (1.48)	0.560	0.554	0.510	0.554	0.542	0.515
	60 (1.78)	0.571	0.567	0.537	0.567	0.563	0.537
	120 (2.08)	0.616	0.616	0.592	0.616	0.631	0.596
	240 (2.38)	0.703	0.703	0.687	0.703	0.715	0.691
	480 (2.68)	0.749	0.749	0.750	0.749	0.772	0.755
	960 (2.98)	0.765	0.765	0.802	0.765	0.818	0.807
6	30 (1.48)	0.525	0.498	0.455	0.499	0.484	0.420
	60 (1.78)	0.518	0.495	0.465	0.495	0.495	0.427
	120 (2.08)	0.553	0.547	0.527	0.547	0.554	0.500
	240 (2.38)	0.622	0.622	0.608	0.622	0.637	0.601
	480 (2.68)	0.708	0.708	0.710	0.708	0.738	0.713
	960 (2.98)	0.721	0.721	0.714	0.721	0.741	0.720
8	30 (1.48)	0.433	0.428	0.388	0.428	0.427	0.336
	60 (1.78)	0.432	0.438	0.383	0.438	0.428	0.334
	120 (2.08)	0.517	0.499	0.433	0.500	0.476	0.400
	240 (2.38)	0.610	0.606	0.572	0.606	0.620	0.564
	480 (2.68)	0.650	0.650	0.644	0.650	0.677	0.637
	960 (2.98)	0.704	0.704	0.780	0.708	0.788	0.772

10	30 (1.48)	0.342	0.339	0.324	0.368	0.354	0.249
	60 (1.78)	0.400	0.378	0.342	0.378	0.370	0.278
	120 (2.08)	0.411	0.422	0.377	0.422	0.406	0.346
	240 (2.38)	0.547	0.541	0.504	0.541	0.540	0.474
	480 (2.68)	0.623	0.622	0.620	0.622	0.655	0.598
	960 (2.98)	0.664	0.664	0.680	0.664	0.704	0.688
12	30 (1.48)	0.314	0.312	0.272	0.312	0.309	0.200
	60 (1.78)	0.352	0.339	0.313	0.340	0.335	0.234
	120 (2.08)	0.405	0.387	0.345	0.387	0.369	0.292
	240 (2.38)	0.476	0.495	0.473	0.496	0.513	0.437
	480 (2.68)	0.611	0.609	0.579	0.609	0.617	0.557
	960 (2.98)	0.640	0.640	0.632	0.640	0.632	0.624

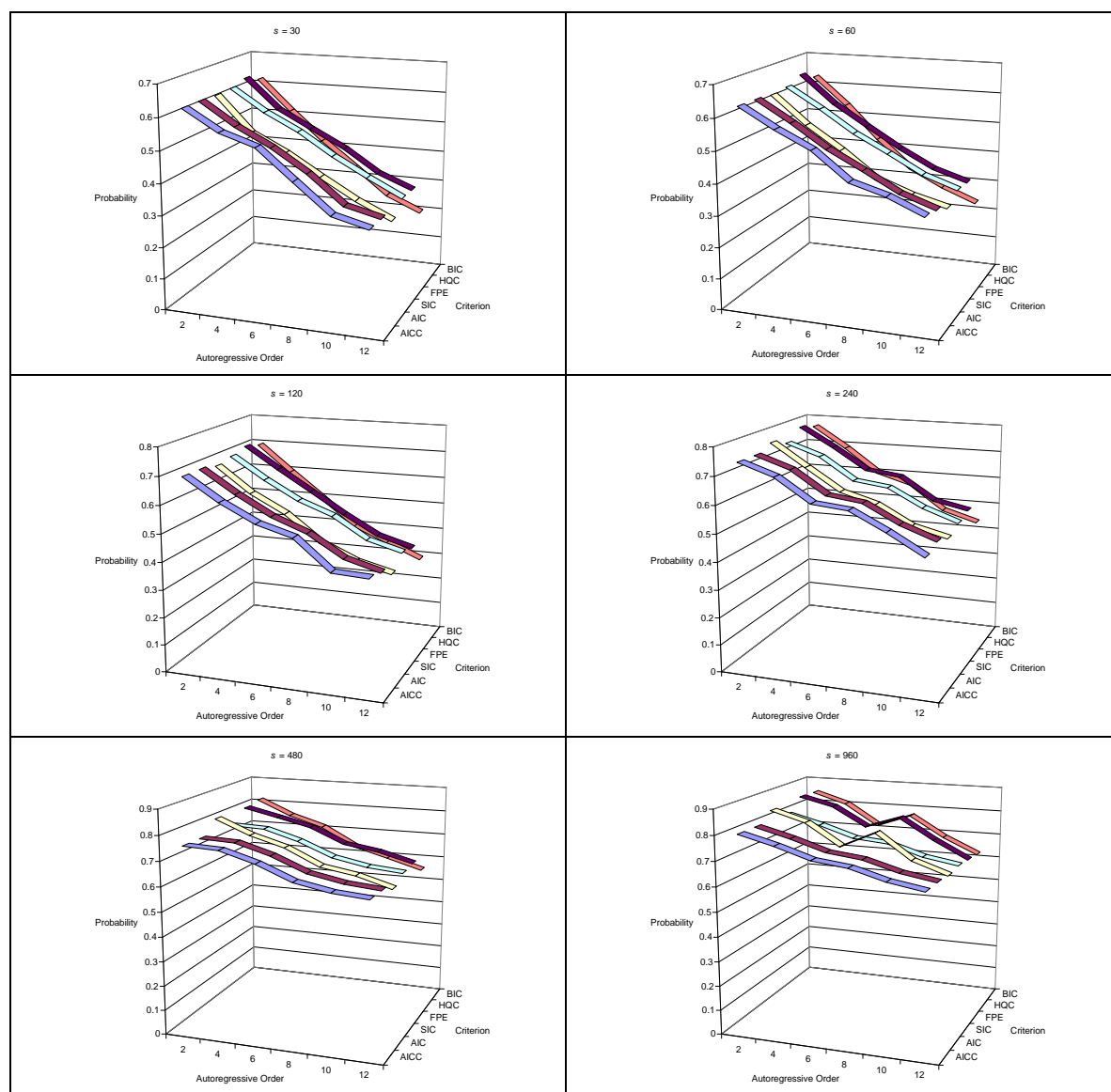


Figure 1. Deterioration in the performance of various criteria in correctly estimated the true AR order as it grows.

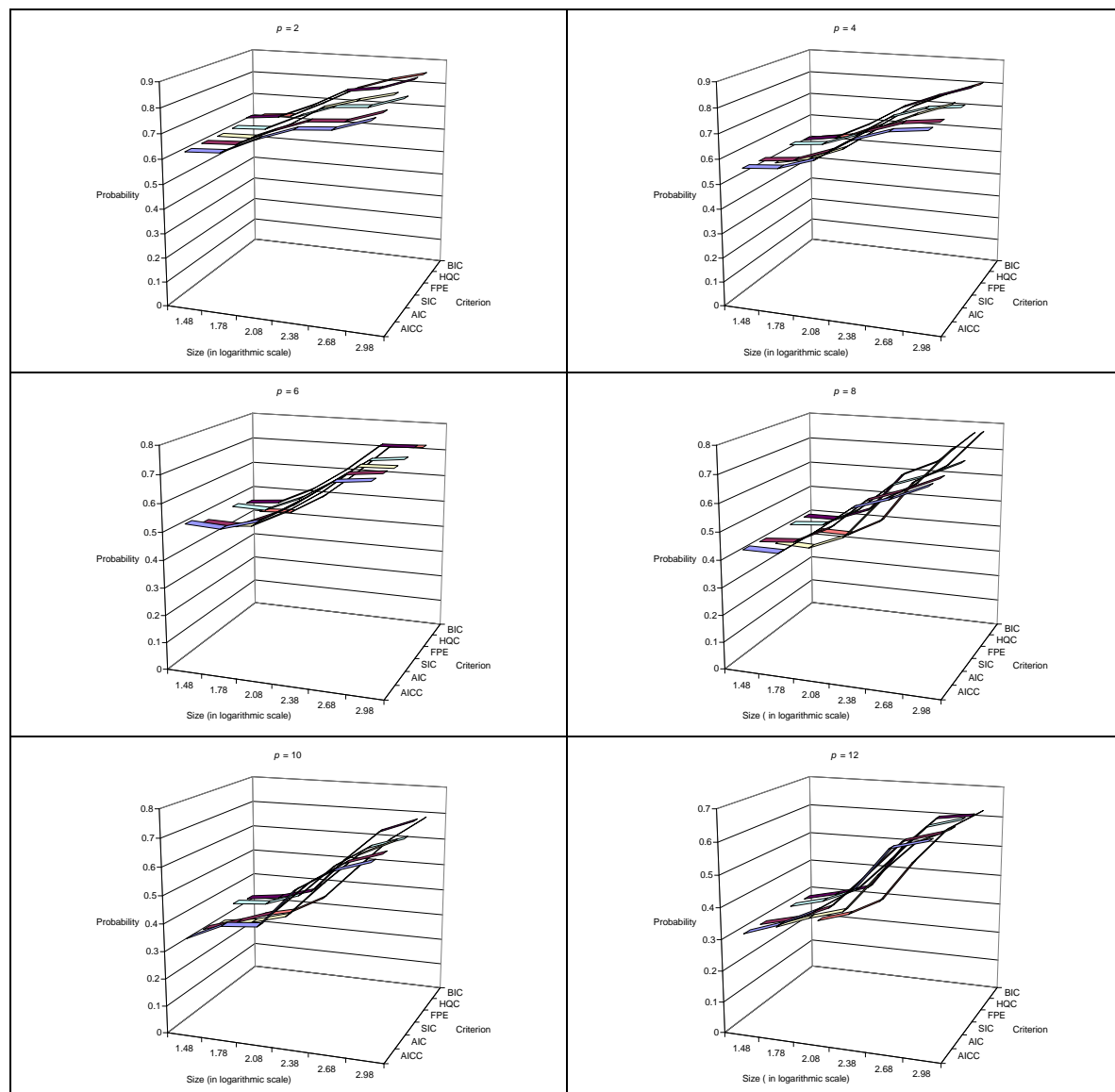


Figure 2. Improvement in the performance of various criteria in correctly estimated the true AR order as sample size grows.

Few other interesting upshots are depicted in Table 1. First, AICC generally perform better than or at least equal to (for relatively large sample size) AIC for all AR order. This empirical evidence is justified if one recalls the fact that the former is a bias-corrected version of the latter Hurvich and Tsai (1989) and therefore the improvement is rational. Besides, it is observed that AICC and AIC have identical accuracy in terms of correctly identified the true order for relatively large sample sizes. This is in line with asymptotic theory, as the penalty factors  $2p$  for AIC (Equation 2) and  $2Tp / (T - p)$  for AICC (Equation 3) are asymptotically equivalent as  $T$  tends to infinity.

Second, it is also shown in Table 1 that AIC and FPE exhibit identical performance regardless of sample size. This finding is rather surprising, as though AIC is the improve version of FPE in terms of consistency Akaike (1973), the performance of the former in terms of accuracy in identifying the true AR order remains unchanged.

An interesting question in mind is whether we can identify from Table 1 the best criterion in selecting the AR order. However, it is difficult to just from Table 1 regarding this matter, as no

criterion is found to consistently perform better than the rest in all cases. Nonetheless, by looking at the average performance across all AR orders as shown in Table 2, it is obvious that AICC dominates the rest for relatively small sample ( $s \leq 120$ ), whereas HQC rears its head for relatively large sample ( $s \geq 240$ ). In fact, an interpolation on the performance of AICC, AIC and HQC as shown in Figure 3 reveals that HQC overtakes AICC and AIC at  $s = 150$ . As such, the choice of criterion should be based on the constraint of sample size one faces and unless one has sample size larger than 150 observations at hands<sup>7</sup>, it is advisable to use AICC to determine the AR order of time series under study. However, the usage of AICC is usually inconvenient as most commercial econometric software do not automatically report this value<sup>8</sup>. Under this constraint (in the absence of AICC) and provided that we have sample up to 150 observations only, AIC and FPE and HQC are better choices than SIC and BIC.

Further analysis of the distribution of the selected orders is conducted and the results are summarized in Tables 3 and 4. Table 3 shows the probability of under estimation. Two main results are revealed by this table: First, for any given sample size, larger AR order provides more room for the occurrence of under estimation. For instance, for  $s = 30$ , the probability of SIC in under estimating the true order for AR orders 2, 4, 6, 8, 10 and 12 are, respectively, 0.363, 0.473, 0.532, 0.604, 0.669 and 0.720. This tendency is also observed in other criteria as is shown in Figure 4. The increase in the room for under estimation has an impact of decreasing the performance of various criteria in finding the right order, as discussed earlier. Note that these findings have connection with the built-in property of the criteria under study which penalize the selection of higher order, in the name of parsimony.

Second, holding AR order constant, the larger the sample size, the less is the tendency of under estimation. This upshot is also shown in Figure 5. It is noteworthy that even in the case of  $p = 12$ , in which the probability of under estimation is the highest among all AR orders, the problem of under estimation will be reduced greatly from more than 60% for  $s = 30$  to around one-third for  $s = 960$  for all criteria. This suggests that we may minimize the problem of under estimation by increasing our sample size.

Table 2

*Average performance of correctly estimated the true order.*

Sample Size	Average Probability across AR Orders					
	AICC	AIC	SIC	FPE	HQC	BIC
30	0.467	0.459	0.429	0.464	0.459	0.390
60	0.484	0.475	0.444	0.475	0.474	0.407
120	0.533	0.528	0.493	0.528	0.525	0.472
240	0.617	0.619	0.602	0.619	0.635	0.589
480	0.683	0.682	0.685	0.682	0.711	0.679
960	0.716	0.716	0.742	0.717	0.756	0.743



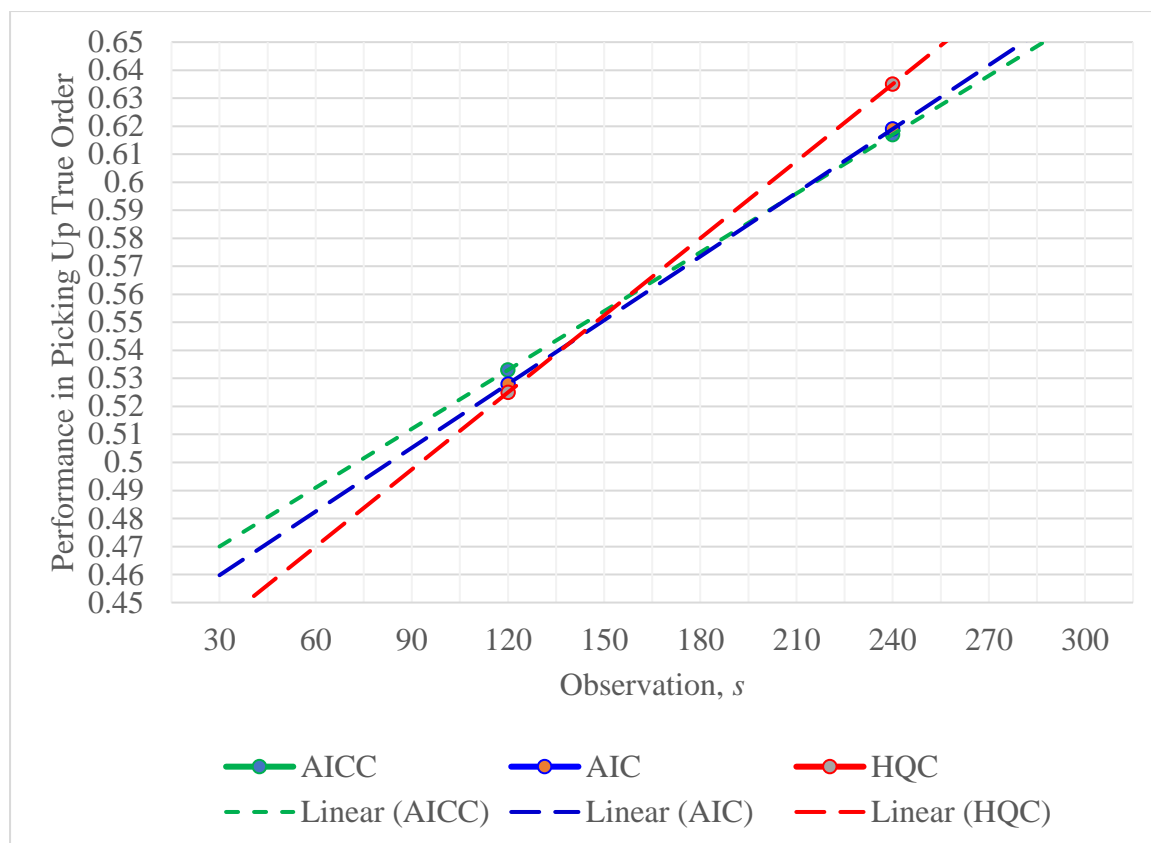


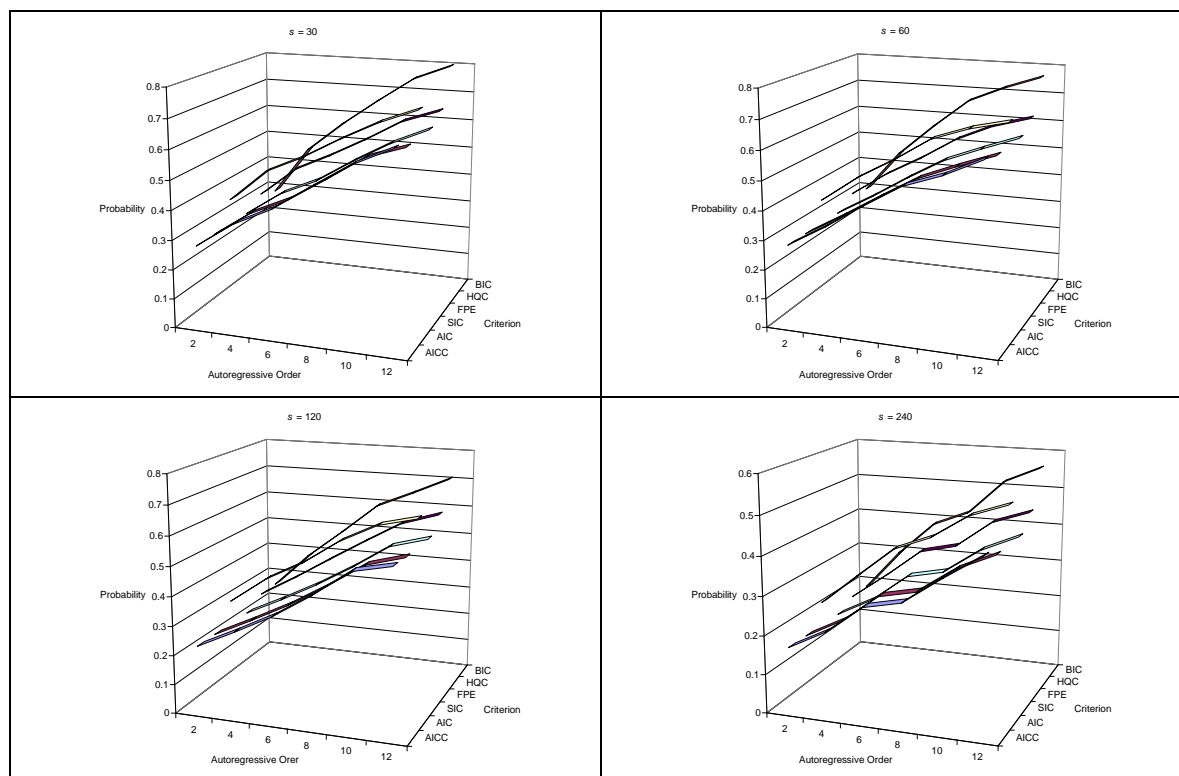
Figure 3: Interpolation of Performance of AICC, AIC and HQC

Table 3

Probability of under estimated the true order of AR process, ( $\hat{p} < p$ ).

True Order $p$	Sample Size (Logarithmic Scale)	Order Selection Criteria					
		AICC	AIC	SIC	FPE	HQC	BIC
2	30 (1.48)	0.276	0.276	0.363	0.276	0.315	0.293
	60 (1.78)	0.280	0.280	0.363	0.280	0.318	0.303
	120 (2.08)	0.227	0.227	0.308	0.227	0.262	0.266
	240 (2.38)	0.166	0.166	0.227	0.166	0.189	0.192
	480 (2.68)	0.133	0.133	0.189	0.133	0.160	0.159
	960 (2.98)	0.090	0.090	0.154	0.090	0.120	0.129
4	30 (1.48)	0.363	0.362	0.473	0.362	0.418	0.463
	60 (1.78)	0.354	0.353	0.453	0.353	0.402	0.451
	120 (2.08)	0.289	0.289	0.399	0.289	0.336	0.387
	240 (2.38)	0.216	0.216	0.307	0.216	0.258	0.299
	480 (2.68)	0.168	0.168	0.247	0.168	0.201	0.234
	960 (2.98)	0.128	0.128	0.192	0.128	0.151	0.182
6	30 (1.48)	0.425	0.420	0.532	0.420	0.479	0.568
	60 (1.78)	0.436	0.431	0.525	0.432	0.472	0.562
	120 (2.08)	0.355	0.355	0.466	0.356	0.415	0.488
	240 (2.38)	0.289	0.289	0.385	0.289	0.333	0.386
	480 (2.68)	0.197	0.197	0.280	0.197	0.235	0.278

	960 (2.98)	0.177	0.177	0.279	0.177	0.227	0.273
8	30 (1.48)	0.509	0.500	0.604	0.500	0.547	0.661
	60 (1.78)	0.513	0.505	0.607	0.505	0.551	0.664
	120 (2.08)	0.436	0.431	0.552	0.431	0.492	0.593
	240 (2.38)	0.310	0.310	0.423	0.310	0.356	0.426
	480 (2.68)	0.248	0.248	0.345	0.248	0.286	0.357
	960 (2.98)	0.148	0.148	0.220	0.148	0.180	0.220
10	30 (1.48)	0.591	0.584	0.669	0.572	0.621	0.748
	60 (1.78)	0.559	0.556	0.652	0.556	0.607	0.719
	120 (2.08)	0.527	0.515	0.616	0.515	0.566	0.649
	240 (2.38)	0.378	0.376	0.485	0.376	0.428	0.517
	480 (2.68)	0.275	0.275	0.378	0.275	0.321	0.400
	960 (2.98)	0.240	0.240	0.320	0.240	0.264	0.304
12	30 (1.48)	0.652	0.627	0.720	0.629	0.667	0.800
	60 (1.78)	0.620	0.603	0.682	0.603	0.644	0.764
	120 (2.08)	0.555	0.551	0.647	0.551	0.607	0.707
	240 (2.38)	0.440	0.419	0.519	0.419	0.461	0.560
	480 (2.68)	0.285	0.285	0.411	0.285	0.346	0.437
	960 (2.98)	0.272	0.272	0.352	0.272	0.328	0.368



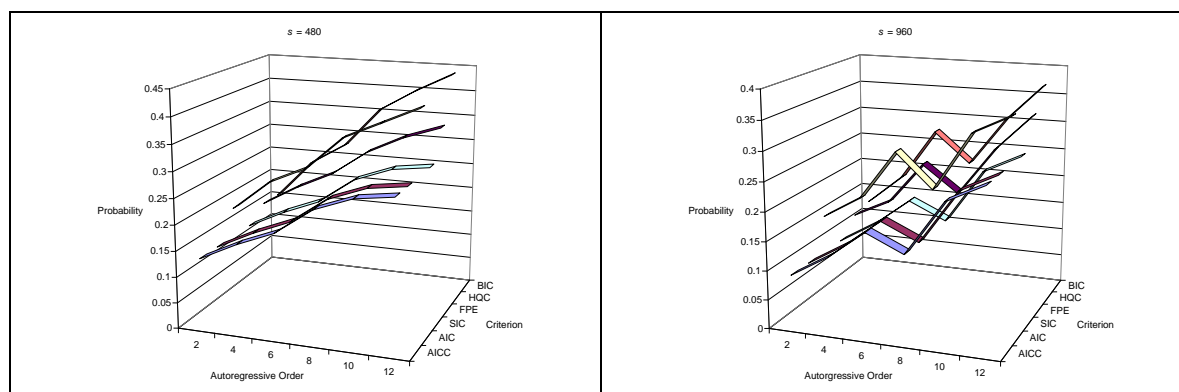


Figure 4. Probability of under estimation increases as AR order grows.

Table 4

Probability of overestimating the true order of AR process, ( $\hat{p} > p$ ).

True Order $p$	Sample Size (Logarithmic Scale)	Order Selection Criteria					
		AICC	AIC	SIC	FPE	HQC	BIC
2	30 (1.48)	0.099	0.099	0.015	0.099	0.045	0.089
	60 (1.78)	0.087	0.088	0.011	0.088	0.028	0.068
	120 (2.08)	0.077	0.077	0.009	0.077	0.026	0.046
	240 (2.38)	0.089	0.089	0.005	0.089	0.026	0.039
	480 (2.68)	0.112	0.112	0.003	0.112	0.036	0.025
	960 (2.98)	0.109	0.109	0.004	0.109	0.028	0.022
4	30 (1.48)	0.077	0.084	0.017	0.084	0.040	0.022
	60 (1.78)	0.075	0.080	0.010	0.080	0.035	0.012
	120 (2.08)	0.095	0.095	0.009	0.095	0.033	0.017
	240 (2.38)	0.081	0.081	0.006	0.081	0.027	0.010
	480 (2.68)	0.083	0.083	0.003	0.083	0.027	0.011
	960 (2.98)	0.107	0.107	0.006	0.107	0.031	0.011
6	30 (1.48)	0.050	0.082	0.013	0.081	0.037	0.012
	60 (1.78)	0.046	0.074	0.010	0.073	0.033	0.011
	120 (2.08)	0.092	0.098	0.007	0.097	0.031	0.012
	240 (2.38)	0.089	0.089	0.007	0.089	0.030	0.013
	480 (2.68)	0.095	0.095	0.010	0.095	0.027	0.009
	960 (2.98)	0.102	0.102	0.007	0.102	0.032	0.007
8	30 (1.48)	0.058	0.072	0.008	0.072	0.026	0.003
	60 (1.78)	0.055	0.057	0.010	0.057	0.021	0.002
	120 (2.08)	0.047	0.070	0.015	0.069	0.032	0.007
	240 (2.38)	0.080	0.084	0.005	0.084	0.024	0.010
	480 (2.68)	0.102	0.102	0.011	0.102	0.037	0.006
	960 (2.98)	0.148	0.148	0.000	0.144	0.032	0.008
10	30 (1.48)	0.067	0.077	0.007	0.060	0.025	0.003
	60 (1.78)	0.041	0.066	0.006	0.066	0.023	0.003
	120 (2.08)	0.062	0.063	0.007	0.063	0.028	0.005

	240 (2.38)	0.075	0.083	0.011	0.083	0.032	0.009
	480 (2.68)	0.102	0.103	0.002	0.103	0.024	0.002
	960 (2.98)	0.096	0.096	0.000	0.096	0.032	0.008
12	30 (1.48)	0.034	0.061	0.008	0.059	0.024	0.000
	60 (1.78)	0.028	0.058	0.005	0.057	0.021	0.002
	120 (2.08)	0.040	0.062	0.008	0.062	0.024	0.001
	240 (2.38)	0.084	0.086	0.008	0.085	0.026	0.003
	480 (2.68)	0.104	0.106	0.010	0.106	0.037	0.006
	960 (2.98)	0.088	0.088	0.016	0.088	0.040	0.008

Another interesting feature depicted in Table 3 is that SIC seems to under estimate the true order more often than AIC. This remark is true for all sample size across all AR orders. This finding from a combination of various sample sizes and a wide range of autoregressive parameters provides formal verification for the claim that SIC tends to underestimate the order than AIC (Baum *et al.*, 2001). In fact, the SIC (other than BIC) is among the criteria which has the highest tendency of under estimation. On the other hand, AICC, AIC and FPE have the lesser tendency of under estimation. However, this does not necessarily imply that the SIC should always be discarded. For instance, if the principle objective is to find a parsimony representation of the behavior of a time series, there is no harm to employ SIC. Moreover, one must not forget the fact that SIC has higher accuracy in identifying the true AR order than for example AIC and AICC for sample size of 480 and more (Table 1). It is also noted remark here that in situation whereby under estimation may induce undesired time series property such as autocorrelation to the estimated results Baum et al (2001); Guerra (2001), AICC, AIC and FPE will be more preferable than the others. This may partially explain why AIC is more preferable in empirical studies<sup>9</sup>. Putting all augments in a sentence, one ought to consider the constraints (e.g. sample size, purpose) one faces in choosing a criterion.

As for Table 4, it shows the results of over estimation. The major information revealed by this table is that over estimation may not be a serious problem in all criteria as the related probability is less than 15% (in most cases, well less than 10%) for all sample size as well as AR order. In other words, if any of the criteria under study fails to identify the true AR order, it is due to under estimation rather than over estimation. As such, we find no empirical evidence supportive of the works of Jones (1975); Shibata (1976) who claim that AIC has the tendency to overestimate the AR order. Nonetheless, the BIC, which is the improve version of AIC Akaike (1979), does in general show slight improvement in reducing the tendency of overestimate, although the related probability for AIC is already very low.

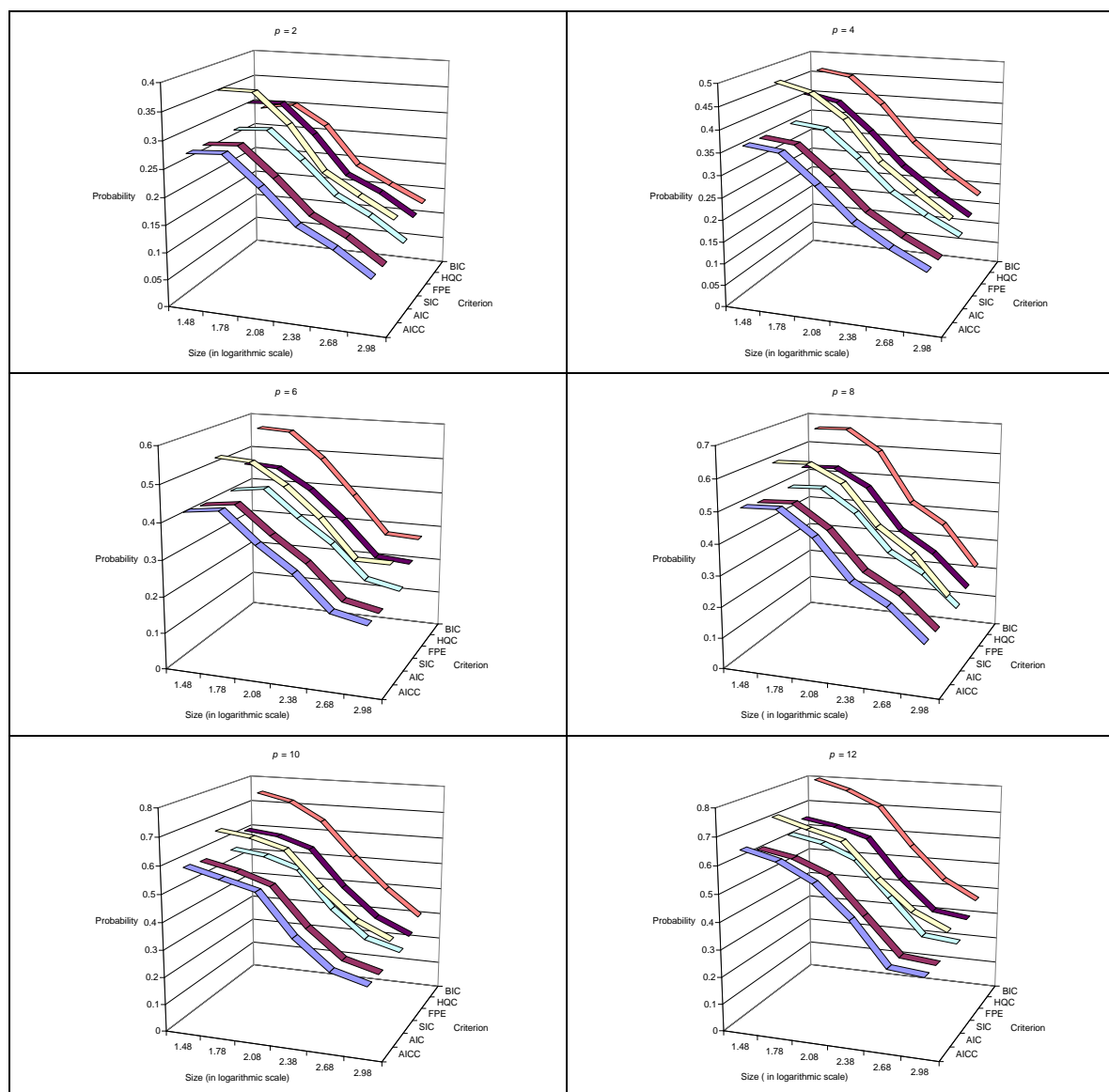


Figure 5. Probability of under estimation decreases as sample size grows.

## Conclusion

The determination of autoregressive order for a time series is especially important in economics studies. Econometric estimation processes including various forms of unit root tests, causality tests, cointegration tests and linearity tests all required the pre-specification of autoregressive order. Various order selection criteria such as the Akaike's bias-corrected information criterion (AICC), Akaike's information criterion (AIC), Schwarz information criterion (SIC), Hannan-Quinn criterion (HQC), final prediction error (FPE) and Bayesian information criterion (BIC) have been employed for this while by researchers in this respect. As the outcomes of these criteria may influence the ultimate findings of a study, a throughout understanding on the empirical performance of these criteria is warranted. This simulation study is specially conducted to shed light on this matter.

The current study independently simulates autoregressive process of various known orders ranging from 2 to 12, for various sample sizes ranging from 30 to 960 observations. 1000 series are generated for each combination of autoregressive order and sample size. Various order selection criteria are then allowed to independently estimate the autoregressive order

for each simulated series, with each criterion yielding some 1000 selected orders for each combination of autoregressive order and sample size. Based on these selected orders, we compute the probabilities in which the true order is correctly identified, under estimate and over estimate. The results, which provide useful insights for empirical researchers are summarized as follows.

Among others, results of study show that the criteria under study manage to pick up the right order well above 60% of the time for low autoregressive order and small sample. However, as autoregressive order increases, their ability to correctly identify the right order decreases as higher orders provide larger room for under estimation. This conclusion is drawn from the facts that the performance of correct estimation drops to around 30% for order 12 in small sample (Table 1) and that the probability of under estimation increases from around 0.3 in the case of order 2 and 30 observations to about 0.7 in the case of order 12 with the same sample size (Table 3). However, the deterioration in performance of these criteria due to large order effect may be surmounted using larger sample, see Figure 5.

Besides, our study demonstrates that AICC performs the best for sample of size 150 and below, whereas HQC turns out to be the best for larger sample. AICC has been shown to improve slightly over AIC and FPE as recommended in Liew (2004). However, researchers usually face the constraints of small sample and the limited availability of AICC (which improves slightly over AIC and FPE) statistics provided by econometric software. In the absence of AICC, results of study across various lag orders suggest that AIC, FPE will still be better choices than other criteria for sample up to the size of 150. It is also noteworthy that although SIC and BIC have the largest tendency to under estimate the autoregressive order, they may be useful if in cases whereby a parsimony order, rather than true order is of interest. On the other hand, in situation whereby under estimation may induce undesired time series property such as autocorrelation to the estimated results AICC, AIC and FPE will be more preferable than the others.

To conclude, via a thorough investigation on the performance of few commonly applied order selection criteria in the presence of various autoregressive orders and sample sizes, this study finds that the choice of order selection criteria should principally be based on the availability of sample size, and the purpose of study. For instance, if our sample size does not exceed 150 observations, Akaike's information criterion family (AICC, AIC) and final prediction error (FPE) are recommended as they are superior to other criteria, while Hannan-Quinn criterion (HQC) performs better than others for sample size of 150 and above. On the other aspects, Schwarz information criterion (SIC), and Bayesian information criterion (BIC) could be useful in cases whereby a parsimony order, rather than true order is of interest; and last but not least, Akaike's information criterion (AIC) and final prediction error (FPE) are more favored to avoid autocorrelation in our ultimate results. The findings in this simulation study may be taken as practical guidelines for future research in the use of order selection criteria in determining the autoregressive order.

### **Acknowledgement**

The financial assistant provided by Universiti Malaysia Sarawak is greatly appreciated.

## References

- Ahmed, M. (2000). Money –income and money-price-causality in selected SAARC countries: some econometric exercises. *Indian Economic Journal*, 48, 55 – 62.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory, B. N. Petrov and F. Csaki (eds.), Akademiai Kiado, Budapest, 267 – 281.
- Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of Autoregressive model fitting. *Biometrika*, 66, 237 – 242.
- Baharumshah, A. Z., Liew, V. K. S., & Lau, E. (2007). Assessing the forecastability of ESTAR model: Some Evidence from ringgit/yen Rate. *ICFAI Journal of Applied Economics*, 1, 7-19.
- Baum, C. F., Barkoulas, J. T., & Caglayan, M. (2001). Nonlinear adjustment to purposing power parity in the post-Bretton Woods era. *Journal of International Money and Finance*. 20, 379 – 399.
- Brockwell, P. J., Davis, R. A. (1994). *ITSM for Windows*. NY: Springer.
- Brockwell, P. J., Davis, R. A. (1996). *Introduction to Time Series and Forecasting*. Springer, New York.
- Caliñgo, A. R., Sison, A. M., Tanguilig III, B. T. (2016). Prediction model of the stock market index using twitter sentiment analysis. *Information Technology and Computer Science*10, 11-21.
- Fainstein, G., and Novikov, I. (2011). The comparative analysis of credit risk determinants in the banking sector of the Baltic States. *Review of Economics & Finance*, 1, 20 – 45.
- Hatmanu, M., Cautisanu, C., & Ifrim, M. (2020). The impact of interest rate, exchange rate and European business climate on economic growth in Romania: An ARDL approach with structural breaks. *Sustainability*, 12, 2798, 1 – 23.
- Hundie, S. K. (2014). Saving, investment and economic growth in Ethiopia: Evidence from ARDL approach to co-integration and TYDL Granger –causality tests. *Journal of Economics and International Finance*, 6(10), 232 – 248.
- Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection small samples. *Biometrika*, 76, 297 – 307.
- Jones, R. H. (1975). Fitting autoregression. *Journal of America Statistics Association*. 70, 590 – 592.
- Leong, C.-M., Puah, C.-H., & Liew, V. K. S. (2018). The impact of divisia money on monetary model of exchange rate in Indonesia. *Economics and Sociology*, 11(2), 52-63.
- Liew, V. K. S. (2000). The performance of AICC as order determination criterion in the selection of ARMA time series models. (Unpublished thesis). Department of Mathematics, Universiti Putra Malaysia, Malaysia.
- Liew, V. K. S., Baharumshah, A. Z., & Lau, E. (2004). Nonlinear adjustment towards purchasing power parity in asean exchange rates. *ICFAI Journal of Applied Economics*, III(6), 7 – 18.
- Liew, V. K. S. (2004). Which lag length selection criteria should we employ? *Economics Bulletin*, 3(33), 1 – 9.
- Lim, K. P., & Liew, V. K. S. (2006). Non-linear mean reversion in stock prices: evidence from Asian markets. *Applied Financial Economics Letters*, 3, 25 – 29.
- Liew, V. K. S., Chia, R. C. J., & Puah, C. H. (2012). Does hysteresis in unemployment occur in OECD countries? Evidence from parametric and non-parametric panel unit roots tests. *International Journal of Economics and Management*, 6(2), 245-257.

- Ng, S., & Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 2001, 69, 1519–1544.
- Olaniyi, E., & Alenoghena, R. O. (2017). Financial Inclusion and gdp per capita in Africa: A Bayesian VAR Model. *Journal of Economics and Sustainable Development* , 8(18), 44 – 57.
- Pickson, R. B., Enning, K. D., & Siaw, A. (2017). Savings-growth nexus in Ghana: Cointegration and causal relationship analyses. *Theoretical Economics Letters*, 7, 139-153.
- Sarantis, N. (1999). Modelling non-linearities in real effective exchange rates, *Journal of International Money and Finance*, 18, 27 – 45.
- Sarantis, N. (2001). Nonlinearities, cyclical behaviour and predictability in stock markets: international evidence. *International Journal of Forecasting*, 17, 439 – 482.
- Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's Information criterion. *Biometrika*, 63, 117 – 126.
- Seo, W., & Kim, L. Y. (2020). Investigating the ripple effect through the relationship between housing markets and residential migration in Seoul, South Korea. *Sustainability* 12, 1225, 1 to 22.
- Tan, H. B., & Baharumshah, A. Z. (1999). Dynamic causal chain of money, output, interest rate and prices in Malaysia: evidence based on vector error-correction modeling analysis, *International Economic Journal*, 13, 103 – 120.
- Tiwari, A. K. (2011). A structural VAR analysis of renewable energy consumption, real GDP and CO2 emissions: Evidence from India", *Economics Bulletin*, 31(2), 1793-1806.
- Xu, Z. (2003). Purchasing power parity, price indices, and exchange rate forecasts. *Journal of International Money and Finance*, 22, 105 – 130.
- Yamada, H. (2000). M2 demand relation and effective exchange rate in Japan: a cointegration analysis. *Applied Economics Letters*, 7, 229 – 232.
- Yang, C., Chen, Y., Niu, L., & Li, Q. (2014). Cointegration analysis and influence rank—A network approach to global stock markets. *Physical A*, 400, 168 – 185.